WORKSHOP: STATISTICS – INTRODUCING SPSS IN ONE HOUR

for Semester Two, 2015

Stats at SLS

Student Learning offers workshops and one-to-one/small group assistance for undergraduate students wanting to improve their statistics skills and understanding of core concepts and topics.

Leila's website for Student Learning Statistics workshop hand-outs & information is here: **www.tinyURL.com/stats-SLS**

Resources for this workshop, including a pdf version of this hand-out and Leila's data files for each problem are available here: **www.tinyURL.com/stats-IS**

Please log in and go here

Students **MUST REGISTER** for each workshop with **Student Learning**

Student Learning free services

Topics we teach in our workshops and provide one-to-one advice on include:

- ✓ **Computer skills**
- ✓ Strategies for succeeding at uni
- ✓ Learning effectively
- ✓ Thinking critically
- ✓ Reading effectively
- ✓ Note-taking
- ✓ Test and Exam skills:
  - o Preparing and revising
  - o Answering multi-choice, essay, and short answer questions
- ✓ Mathematics
- ✓ **Statistics**
- ✓ Writing skills including:
  - o Paraphrasing, summarising, and quoting
  - o Sentence and paragraph writing
  - o Editing and punctuation
  - o Essay writing (analysing the question, planning, introductions, conclusions)
  - o Referencing

- **Te Fale Pouāwhina (TFP) – for Māori and Pasifika students**
  Te Fale Pouāwhina (TFP) provides a friendly and professional service based on kaupapa Māori and Pacific principles. Māori and Pacific Learning Advisers are passionate about Māori and Pacific student success. The focus is on academic skills development at both undergraduate and postgraduate levels. TFP Learning Advisers work with a range of Māori and Pacific services across the University. Face-to-face learning options include workshops, study groups, wānanga/fono and advisory sessions. Contact Matt Tarawa or Mona O'Shea, or go to ***www.tinyURL.com/SLS-TFP***

- **Learning disabilities students**
  Learning assessments are available for students with specific learning disabilities; academic assistance is available for these students and those with mental health impairments.
  If you have any special learning requirements, please feel free to discuss this with Leila in person or via email or contact Student Learning Reception directly by phone on 373-7599 ext. 88850 or email sls@auckland.ac.nz.
  More information is available here: ***tinyURL.com/SLS-LDS***

- **English Language Enrichment (ELE) Language Learning Groups**
  Improve your English with 'Let's talk!' Meet with other students and an ELE Learning Adviser and talk about things that you *want to talk about. No need to enrol. Just come. Mondays and Wednesdays 11-12 and Fridays 3-4 at ELE (Room 101, Level One, Info Commons Building).*
  More information is available here: ***tinyURL.com/SLS-ELE***

- **Language Exchange (LEX)**
  LEX is a non-credit activity at the University of Auckland. It gives you the opportunity to develop a language you want to learn or improve and helps other students to develop a language they are learning. All LEX participants are volunteers and the programme is free.
  More information is available here: ***tinyURL.com/SLS-LEX***

Student Learning location and contact details

Student Learning Reception, City Campus, Room 315.320 (third floor, Kate Edgar Information Commons Building); 09-923 8850; sls@auckland.ac.nz; **www.library.auckland.ac.nz/student-learning**

Statistical help available from Leila Boyle at Student Learning

Each semester, Leila Boyle at Student Learning offers statistical help through a number of workshops and advisory sessions (either one-to-one or with study groups of two or more people) by appointment.

Student Learning Statistics Advisory Sessions

Book your preferred time with Leila here: **www.tinyurl.com/appt-stats**, or contact her directly:

Leila Boyle
Undergraduate Statistics Assistance at Student Learning
Room 315.336 (third floor, Kate Edgar Information Commons Building)
l.boyle@auckland.ac.nz; (09) 923-9045; 021 447-018

Student Learning Statistics Workshops

Workshops are run in a relaxed environment, and allow plenty of time for questions. In fact, this is encouraged ☺

Please make sure you bring your calculator with you to all of these workshops. Either a scientific calculator or a graphics calculator is fine - it should be able to calculate the mean and standard deviation of a set of data.

- **Preparation at the beginning of the semester:**

A couple of preparation workshops are run at the beginning of the semester to get students off to a good start. There are multiple identical sessions of these one-hour long on weekdays over Weeks 1 to 3 of the semester – enrol in whichever one suits your schedule.

  - Basic maths and calculator skills for Statistics **www.tinyURL.com/stats-BM**
  - Finding the sample mean and standard deviation (using a scientific or graphics calculator) **www.tinyurl.com/stats-FS**

- **First half of the semester**

Five theory workshops and one computing workshop are during the first half of the semester:

  - Polls, Surveys, Experiments and Observational Studies **www.tinyURL.com/stats-PS**
  - Basic Data Analysis **www.tinyURL.com.com/stats-BDA**
  - Proportions and Probabilities **www.tinyURL.com/stats-PP**

- Confidence Intervals: *Proportions* **www.tinyURL.com/stats-CIP**
- Confidence Intervals: *Means* **www.tinyURL.com/stats-CIM**
- **Computer Workshop:** Introducing SPSS in One Hour **www.tinyURL.com/stats-IS**

Second half of the semester

Four theory workshops and one computing workshop are during the second half of the semester:

- **Statistics Theory Workshops**
  - Hypothesis Tests: *Proportions* **www.tinyURL.com/stats-HTP**
  - Hypothesis Tests: *Means and medians* **www.tinyURL.com/stats-HTM**
  - Chi-Square Tests **www.tinyURL.com/stats-CST**
  - Regression and Correlation **www.tinyURL.com/stats-RC**
  - **Computer Workshop:** Hypothesis Tests: *SPSS (Means and Medians) in One Hour* **www.tinyURL.com/stats-HTS**

Please enrol in each of your preferred workshops at the Student Learning by either:

- **Going online at:** www.library.auckland.ac.nz/booking **(search for *Leila*)**

OR

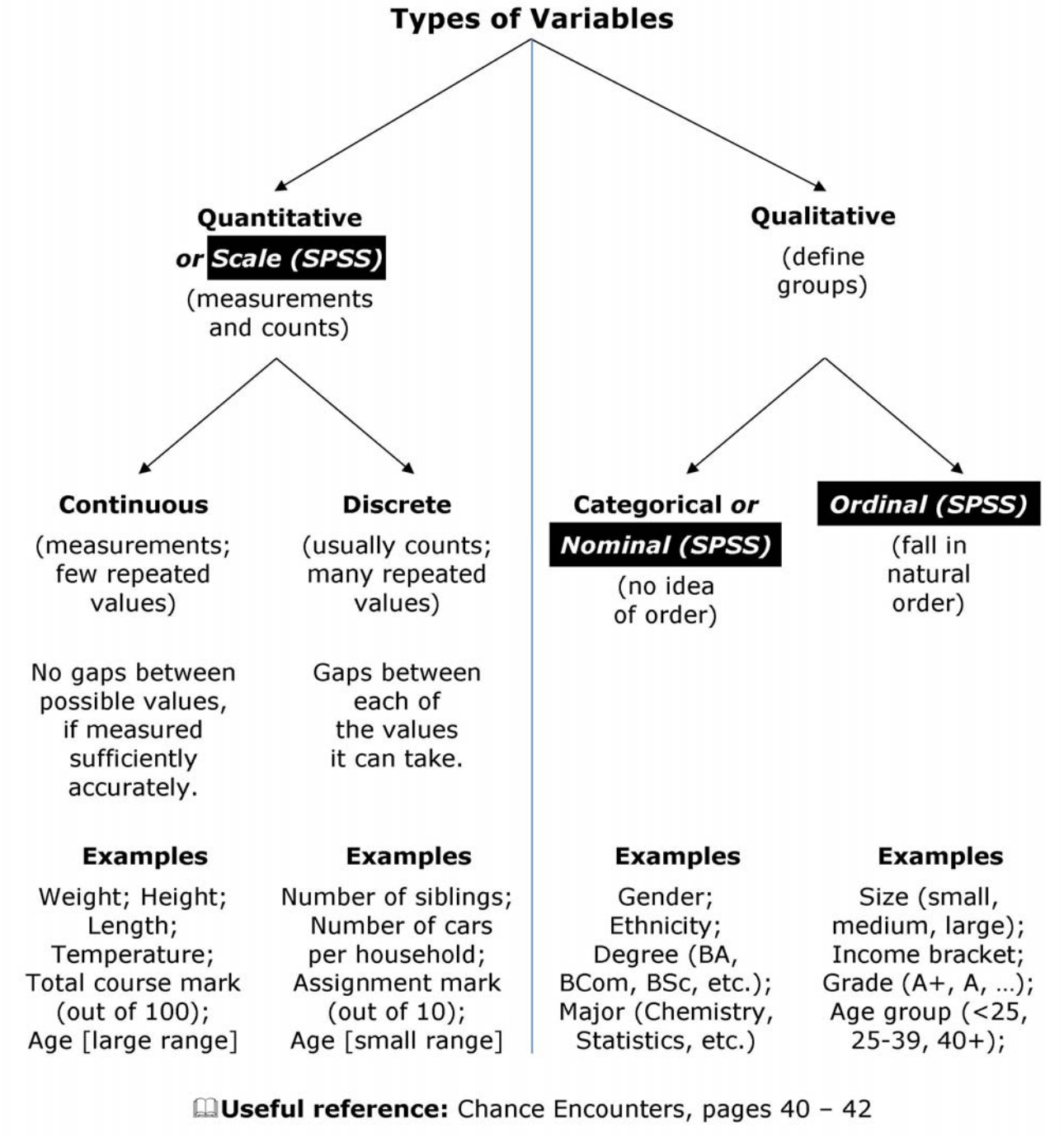
- **Going to Student Learning in person (3rd floor, Kate Edgar Information Commons Building, corner of Alfred and Symonds Streets)**

Useful Websites

- Student Learning webpage: www.library.auckland.ac.nz/student-learning
- For online workshop enrolment go to www.library.auckland.ac.nz/booking
  - Click on ***Undergraduate workshops***, and ***Refine your results*** by clicking on appropriate ***Topics***
  - For statistics workshops, ***Search workshops*** by ***statistics***.
- Cecil: https://cecil.auckland.ac.nz
- **Leila's website for Student Learning Statistics workshop hand-outs and information: www.tinyURL.com/stats-SLS**

Basic Data Analysis

Want to check you understand this material and do some practice questions? **Find out more** at the ***Basic Data Analysis*** workshop, held in the first half of the semester.



📖**Useful reference:** Chance Encounters, pages 40 – 42

Displaying/graphing your data

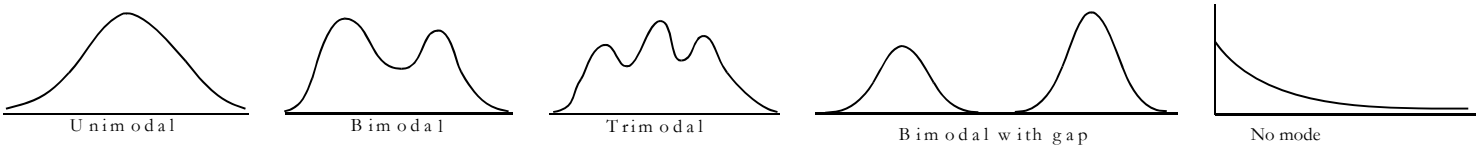
Common displays of data include tables and graphs such as dotplots, stem-and-leaf plots, boxplots, histograms and bar charts. Choosing which of these to use will depend on the type of variable/s you have collected and the relationships you are attempting to explore.

Displaying continuous quantitative variables

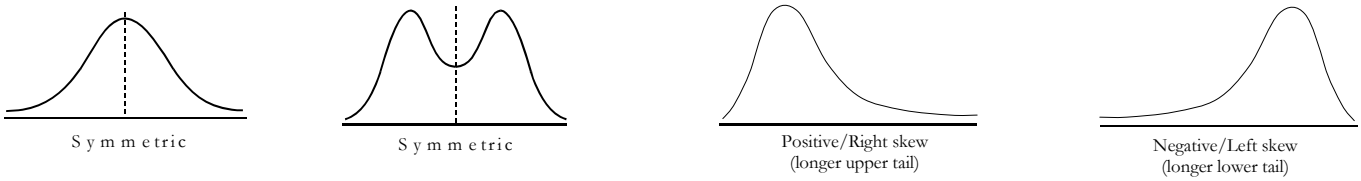
Display tools:

- Appropriate plots for continuous quantitative variables are:
  - Dot plot – small data sets, $n \leq 20$
  - Stem-and-leaf plot – moderate data sets, $15 \leq n \leq 150$
  - Box plot – moderate to large data sets, $n \geq 20$
  - Histogram – large data sets, $n \geq 50$

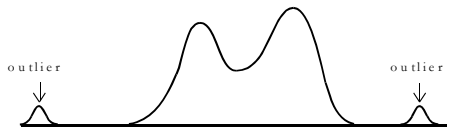
- Features to look for and comment on in the above plots are:
  - Centre and spread
  - Modality – How many modes/peaks does the data have?



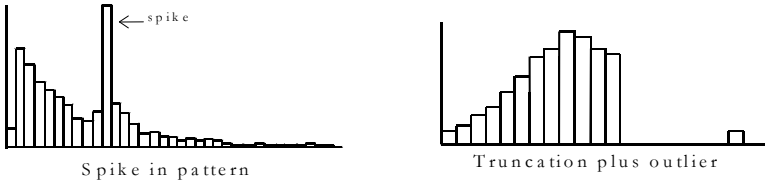
  - Symmetry or skewness – Is the data symmetric or skewed?



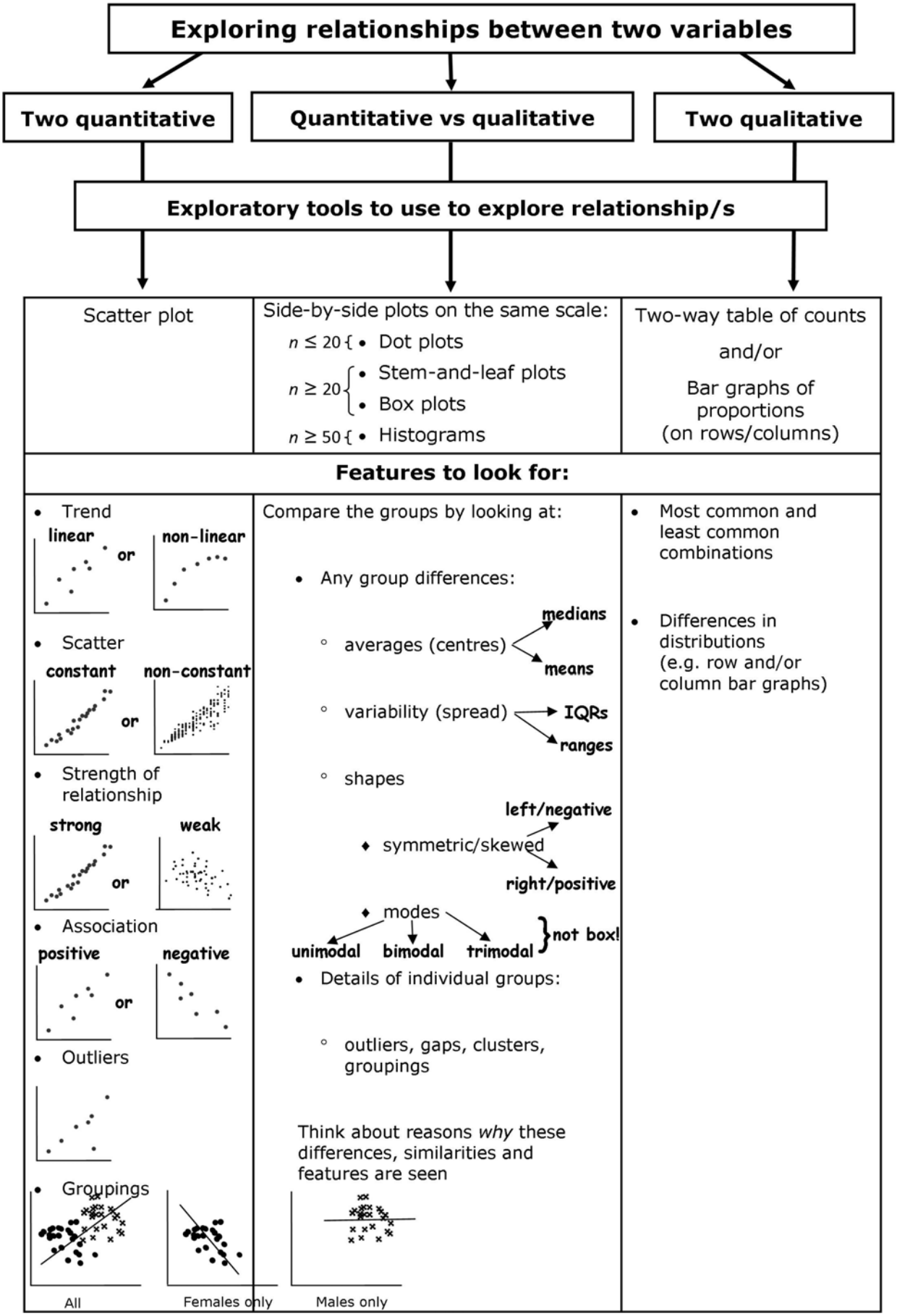
  - Outliers:
    - Are observations which are far from the bulk of the data
    - Search for a reason for their existence
    - Only delete outliers if they are found to be mistakes
    - SPSS can identify outliers for you!



  - Abrupt changes:



📖**Useful reference:** Chance Encounters, pages 58 – 60


Exploring relationships between two variables
Two quantitative
Quantitative vs qualitative
Two qualitative
Exploratory tools to use to explore relationship/s
Scatter plot
Side-by-side plots on the same scale:
n ≤ 20 { • Dot plots
n ≥ 20 { • Stem-and-leaf plots
• Box plots
n ≥ 50 { • Histograms
Two-way table of counts
and/or
Bar graphs of proportions (on rows/columns)
Features to look for:
• Trend
linear
or
non-linear
• Scatter
constant
or
non-constant
• Strength of relationship
strong
or
weak
• Association
positive
or
negative
• Outliers
• Groupings
All
Females only
Males only
Compare the groups by looking at:
• Any group differences:
◦ averages (centres) → medians / means
◦ variability (spread) → IQRs / ranges
◦ shapes
♦ symmetric/skewed → left/negative / right/positive
♦ modes
unimodal
bimodal
trimodal
not box!
• Details of individual groups:
◦ outliers, gaps, clusters, groupings
Think about reasons why these differences, similarities and features are seen
• Most common and least common combinations
• Differences in distributions (e.g. row and/or column bar graphs)

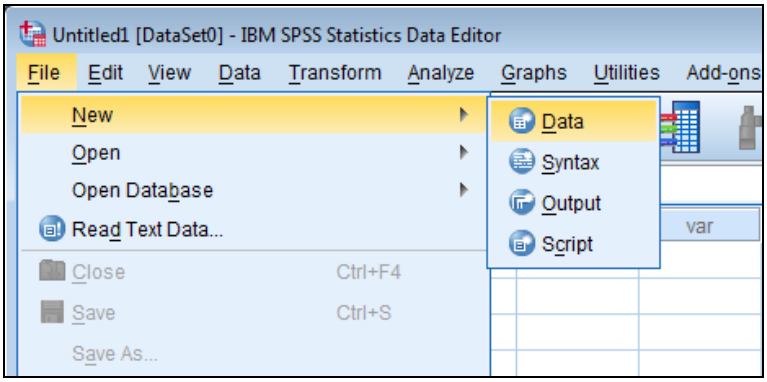
SPSS Basics

You can use SPSS to plot data and perform simple calculations.

Opening a File in SPSS

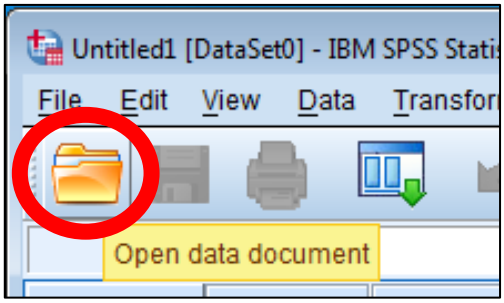
To create a new Data file in SPSS

1. Select **File** → **New** → **Data**.



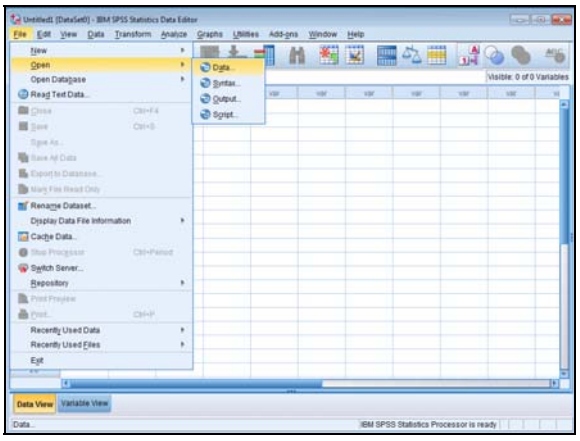
OR

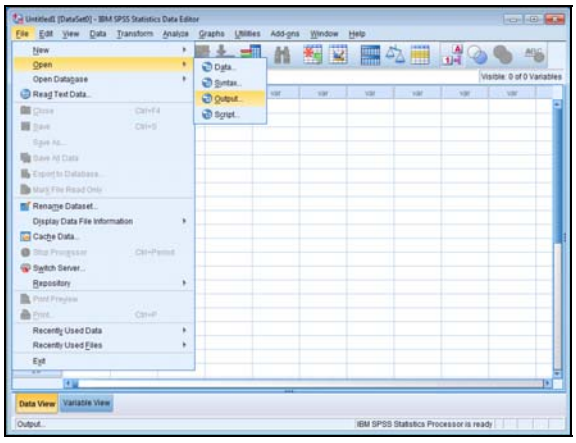
2. Click the **Open data document** icon on the SPSS **Data Editor Toolbar**.



To open an existing file in SPSS

1. Data file: click **File** → **Open** → **Data**.
   Output file: click **File** → **Open** → **Output**.

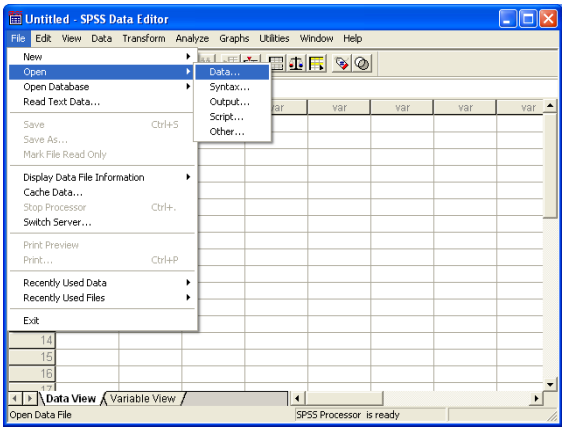




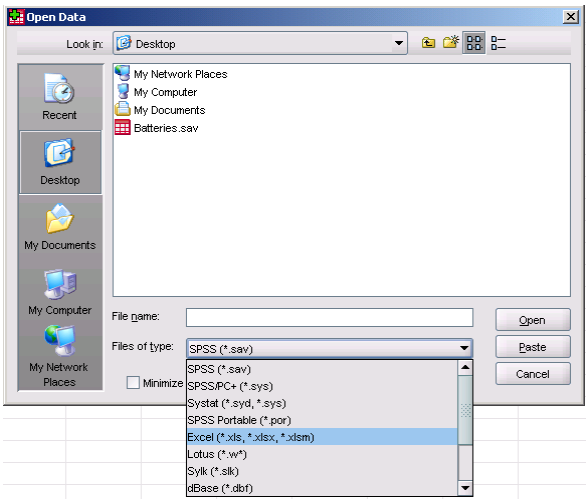
2. Choose the file.
   Find and click the required file and click **Open**.

Importing data from an *Excel* file

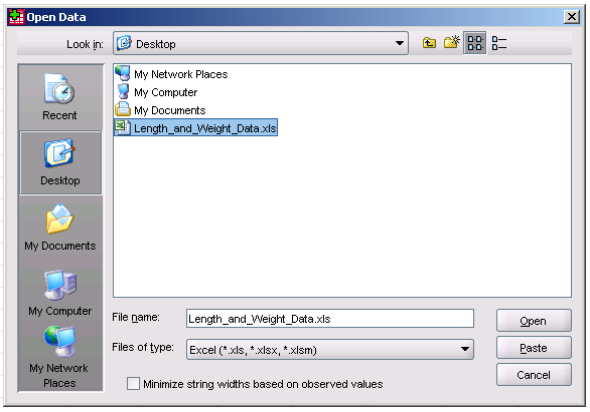
1. Click **File** → **Open** → **Data**.



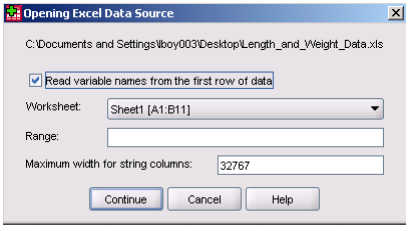
2. Choose the file type.
In the **Files of Type** box, choose **Excel (*.xls)**.



3. Choose the file.
Find and click the required file and click **Open**.



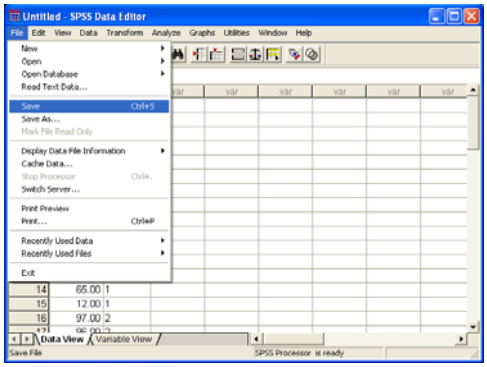
4. Click **Continue/OK**.

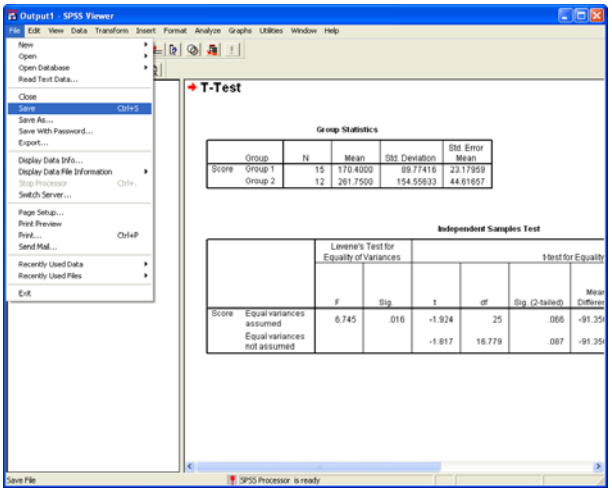


Saving

In SPSS, data and output are saved separately.

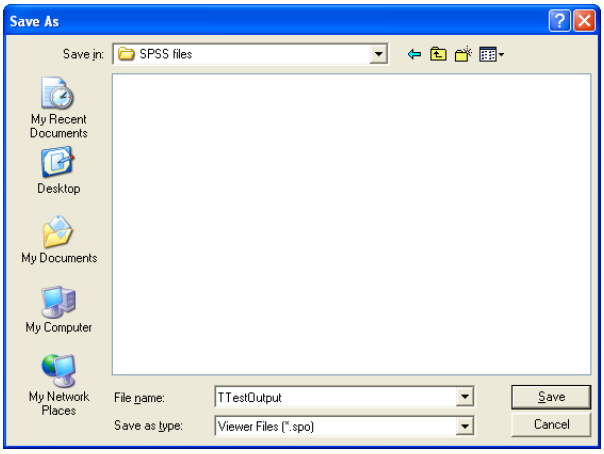
1. Click **File** → **Save**.



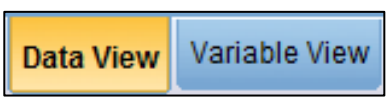


2. Enter the file name.

   Type a name for the file in the **File name** box and click **Save**.



What can we see in the SPSS Data Editor?



In SPSS, there are two different views, the **Data View** and the **Variable View**. The **Data View** and the **Variable View** are two interlinked representations of the data contained in SPSS's **Data Editor**. You can switch between these views by clicking on the tabs with these names at the bottom left of SPSS's **Data Editor** window; whichever tab is orange is the view currently on the screen.

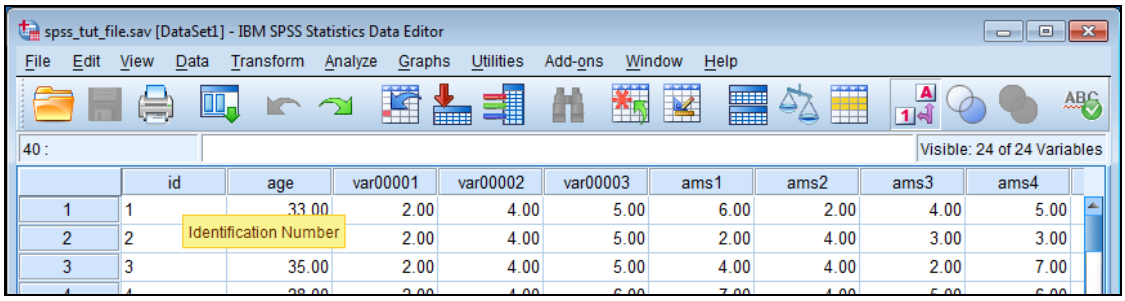
Data View

To view the data view, click the **Data View** tab at the bottom of the window.

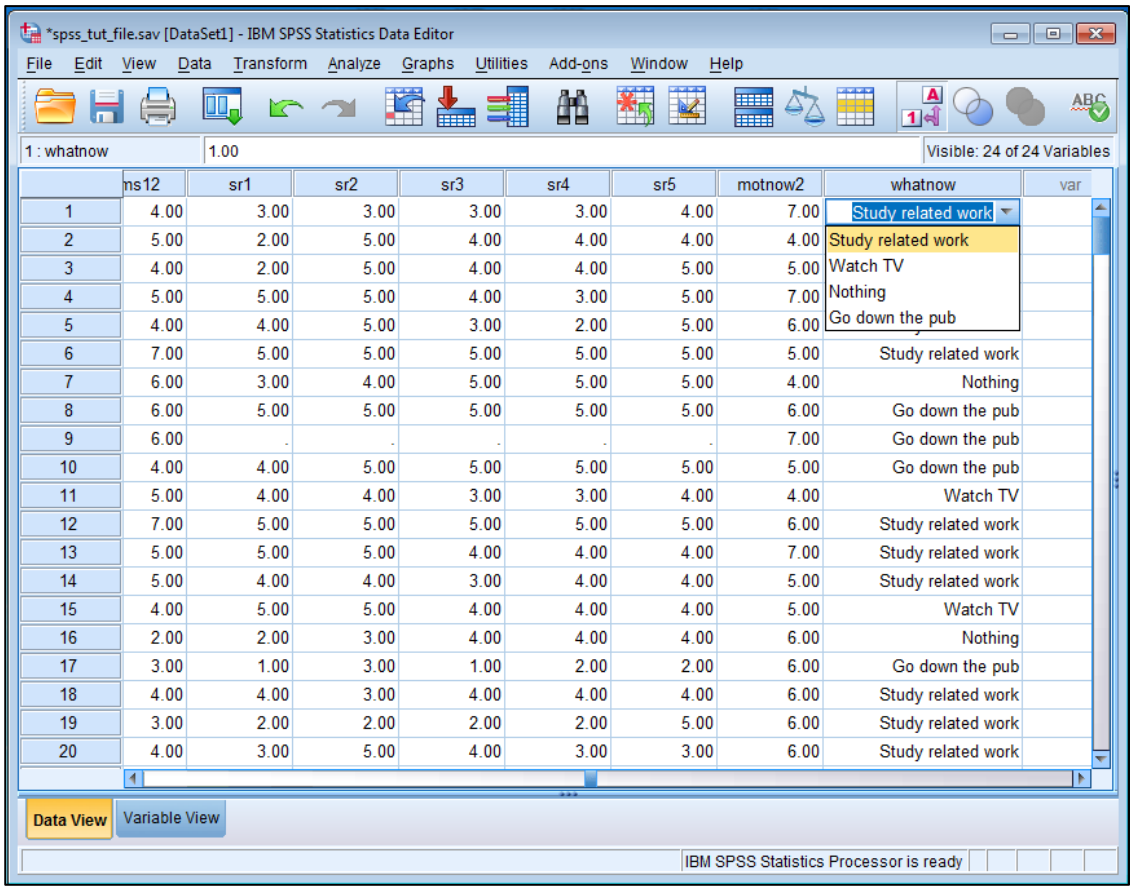
All data is entered in the data view. Each row corresponds to one case (an individual, an experimental unit, etc.), and each column is a variable.

You can also view labels in the data view.

To view variable labels, hold the mouse over the variable names at the top of each column. The variable label will appear.



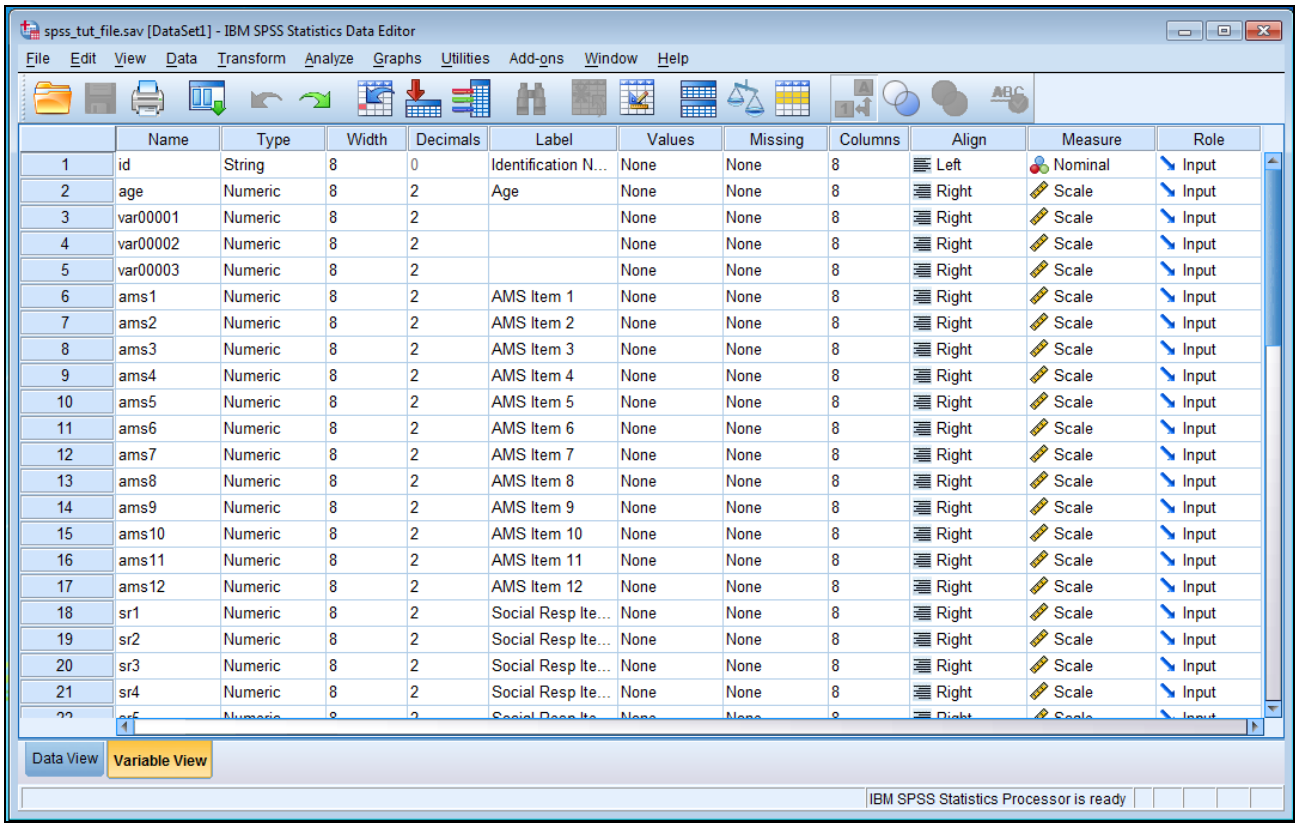
To view value labels, click the **Value Labels** button  . The labels will now appear instead of the values. You can see all of the labels available for any variable by triple clicking on the right side of any cell in that variable's column.



Variable View

To view the variable view, click the **Variable View** tab at the bottom of the window.

Variables are created in the **Variable View** window. Each row corresponds to one variable, and the columns contain information about each variable.



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	String	8	0	Identification N...	None	None	8	Left	Nominal	Input
2	age	Numeric	8	2	Age	None	None	8	Right	Scale	Input
3	var00001	Numeric	8	2		None	None	8	Right	Scale	Input
4	var00002	Numeric	8	2		None	None	8	Right	Scale	Input
5	var00003	Numeric	8	2		None	None	8	Right	Scale	Input
6	ams1	Numeric	8	2	AMS Item 1	None	None	8	Right	Scale	Input
7	ams2	Numeric	8	2	AMS Item 2	None	None	8	Right	Scale	Input
8	ams3	Numeric	8	2	AMS Item 3	None	None	8	Right	Scale	Input
9	ams4	Numeric	8	2	AMS Item 4	None	None	8	Right	Scale	Input
10	ams5	Numeric	8	2	AMS Item 5	None	None	8	Right	Scale	Input
11	ams6	Numeric	8	2	AMS Item 6	None	None	8	Right	Scale	Input
12	ams7	Numeric	8	2	AMS Item 7	None	None	8	Right	Scale	Input
13	ams8	Numeric	8	2	AMS Item 8	None	None	8	Right	Scale	Input
14	ams9	Numeric	8	2	AMS Item 9	None	None	8	Right	Scale	Input
15	ams10	Numeric	8	2	AMS Item 10	None	None	8	Right	Scale	Input
16	ams11	Numeric	8	2	AMS Item 11	None	None	8	Right	Scale	Input
17	ams12	Numeric	8	2	AMS Item 12	None	None	8	Right	Scale	Input
18	sr1	Numeric	8	2	Social Resp Ite...	None	None	8	Right	Scale	Input
19	sr2	Numeric	8	2	Social Resp Ite...	None	None	8	Right	Scale	Input
20	sr3	Numeric	8	2	Social Resp Ite...	None	None	8	Right	Scale	Input
21	sr4	Numeric	8	2	Social Resp Ite...	None	None	8	Right	Scale	Input

Name

This column contains the name of each variable. To enter the name, click in the cell and type the variable name.

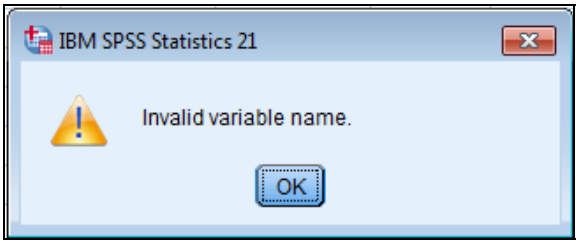
Variable names are abbreviations of what the variable is. There is, the name must start with an upper or lowercase letter and may include any combination of letters, numbers and underscores but no other characters are allowed.

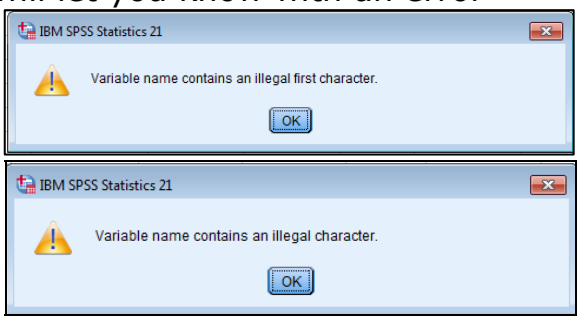
There are rules that govern what we can actually call a variable:

- Each variable name must be unique; duplication is not allowed.
- Variable names must begin with a letter (upper or lower case) or an at sign (@).

- Variable names cannot end with a full stop.
- Variable names cannot contain spaces.
- Depending on the language you are using, variable names cannot exceed 64 characters (in English, French, German, Spanish, Italian, Hebrew, Russian, Greek, Arabic, and Thai, for example) or 32 characters (in Japanese, Chinese, and Korean, for example).
- Variable names may include letters (upper or lower case), numbers and some other characters such as the full stop (.), the at sign (@), the hashtag (#), the dollar sign ($) or the underscore (_). For example, A._$@#1 is a valid (but not particularly meaningful) variable name!
- Reserved keywords cannot be used as variable names: ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH.

Don't bother to learn these rules – SPSS will let you know with an error message if you break one!





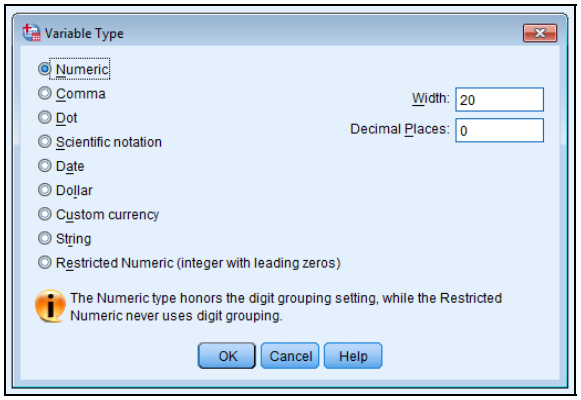
Often people truncate or abbreviate variable names and assign a **Label** which gives a more detailed description of the variable (including units, if appropriate) and/or a more meaningful name.

Type

This contains the type of the variable – usually Numeric. When entering qualitative data, you should always use the Numeric type. You can then use numbers (e.g. enter 1, 2, 3, … as codes) when entering data and use the value labels in the **Value** column to make them more meaningful.

If you want to change the type, click in the cell and then click  .

Choose the type and click **OK**.



Width

If the variable is of type String, this represents the maximum number of characters the string can contain. You should make sure this is high enough to allow all of your data to be entered correctly.

To enter the width of the string, click in the cell and type the width.

Decimals

If the variable is of type Numeric, this represents the number of decimal places the variable can contain.

To enter the number of decimal places, click in the cell and type the number.

Label

Variable labels are used to give a more meaningful description of the variable than the variable name. Understanding the computer output is made easier by using variable labels. They can be up to 256 characters long, and any characters, including spaces, are allowed.

To enter a label, click in the cell and type the label.

Values

Value labels are used to give more meaningful descriptions of the numerical values used for qualitative (grouping/categorical) data.

For example, in our dataset, we have allocated the codes 1 and 2 for the two different levels of the gender group. Understanding the 1s and 2s in the computer output is made easier by using value labels.

They can be of any length, and spaces and all other characters are allowed.

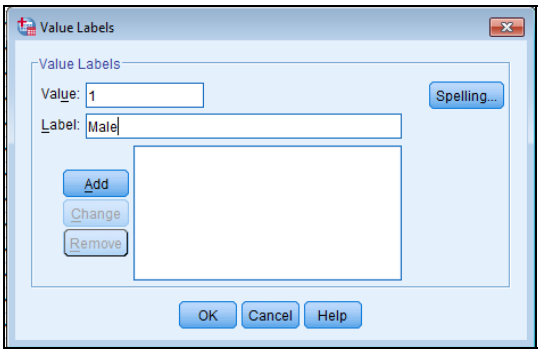
To enter value labels, click in the cell and then click  .

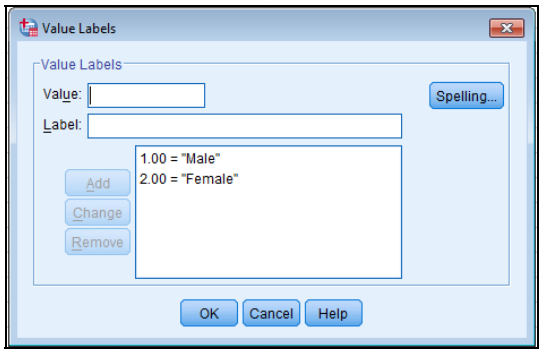
In the **Value** box, type one of the values that your variable takes.

In the **Value Label** box, type the label that you want that value to have.

Click **Add**.

Repeat this process for all values. Then click **OK**.





It is also possible to enter actual words i.e. ***String*** *variables into SPSS, but if we entered* ***Male*** *and* ***Female*** *in this way we would not be able to use them in statistical analyses. If you do wish to enter words as part of your data set however, in the* ***Variable View****, in the* ***Type*** *column for your variable, click on the box with the three dots that appears, and change to* ***String****, then click the* ***OK*** *button.*

Entering data

1. Set up the variables in the **Variable View**.

   Click **Variable View**. Enter each variable that you will be using.

2. Type in the data in the **Data View**.

   Click **Data View**. Enter all the data.

It is good practice to assign a unique identifier to each subject. This will help you find each individual quickly and easily if you find you have any data entry mistakes or if they are identified as an outlier.

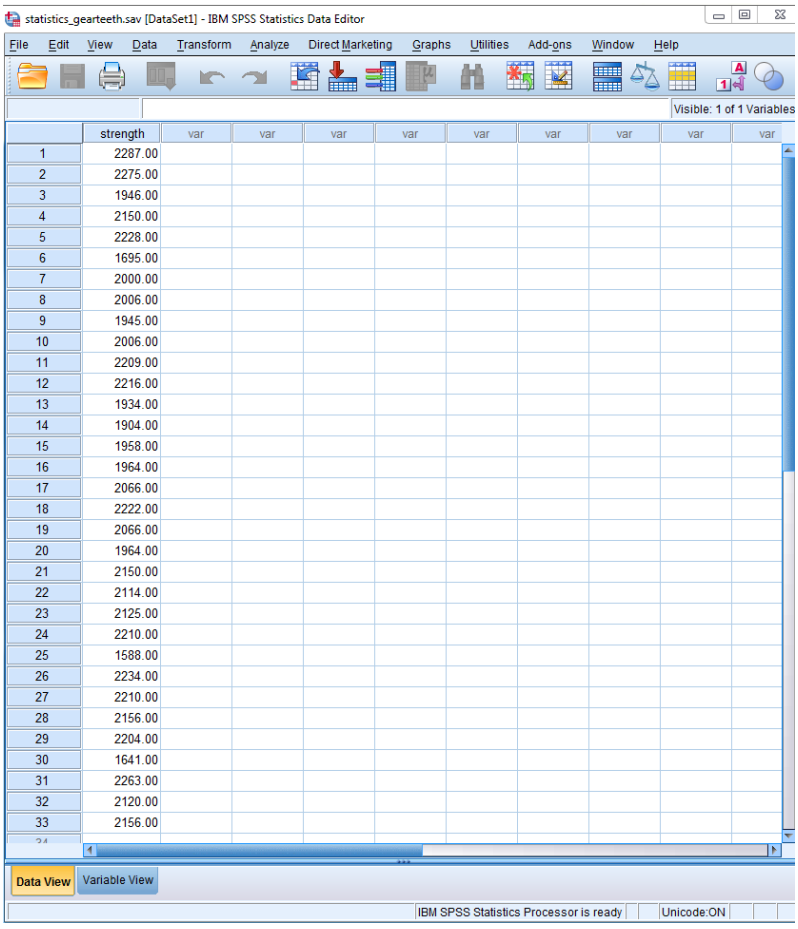
Generating Descriptive Statistics and Creating a Stem and Leaf Plot and a Box Plot in SPSS

Example: Generate descriptive statistics and create a stem-and-leaf plot and a box plot for the breaking strengths of gear teeth in certain positions of a gear.

1. Enter the data into SPSS

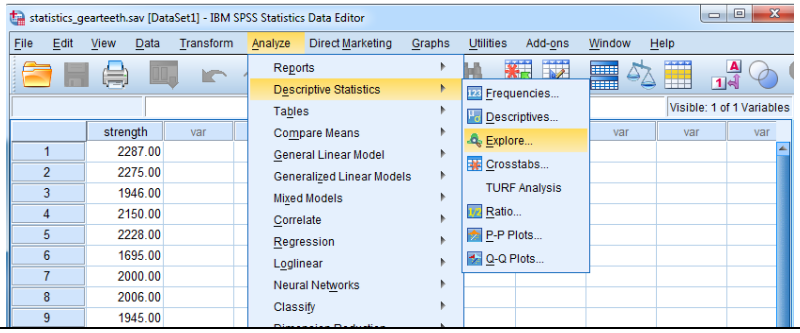
   OR open the GearTeeth.sav file from **www.tinyURL.com/stats-IS**.

   Label **strength** as **Breaking Strengths of Gear Teeth** in the **Variable View**.



2. Choose the analysis tool: **Explore.**

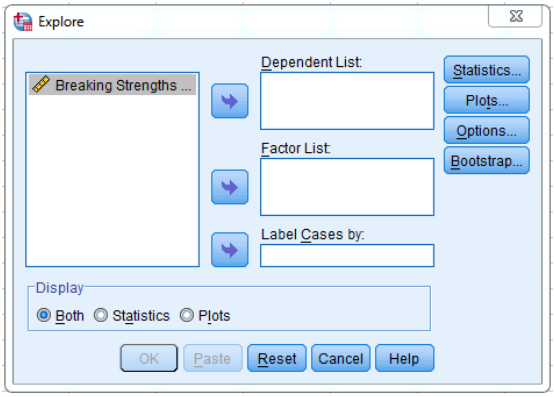
   Click **Analyze** → **Descriptive Statistics** → **Explore**.

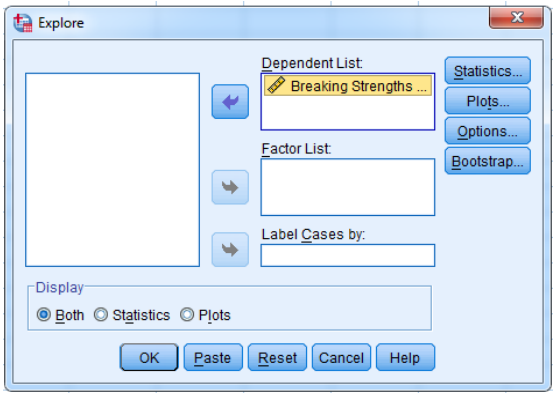


3. Select the relevant variable(s).

   Quantitative variable(s) → **Dependent List** box.

   Click **Breaking Strengths of Gear Teeth [strength]**. Click the first  . Then click **OK**.





4. The results appear in the Output Window.

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Breaking Strengths of Gear Teeth	33	100.0%	0	0.0%	33	100.0%

Descriptives

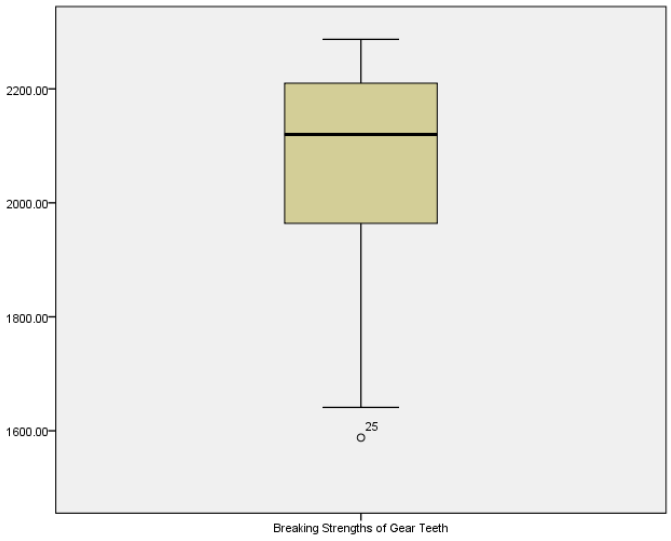
			Statistic	Std. Error
Breaking Strengths of Gear Teeth	Mean		2067.0303	31.01907
	95% Confidence Interval for Mean	Lower Bound	2003.8465	
		Upper Bound	2130.2141	
	5% Trimmed Mean		2080.5253	
	Median		2120.0000	
	Variance		31752.030	
	Std. Deviation		178.19099	
	Minimum		1588.00	
	Maximum		2287.00	
	Range		699.00	
	Interquartile Range		249.00	
	Skewness		-1.152	.409
	Kurtosis		1.060	.798

Breaking Strengths of Gear Teeth

```
Breaking Strengths of Gear Teeth Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     1.00 Extremes    (=<1588)
     2.00       16 .  49
      .00       17 .
      .00       18 .
     7.00       19 .  0344566
     5.00       20 .  00066
     7.00       21 .  1225555
    11.00       22 .  00111223678
```



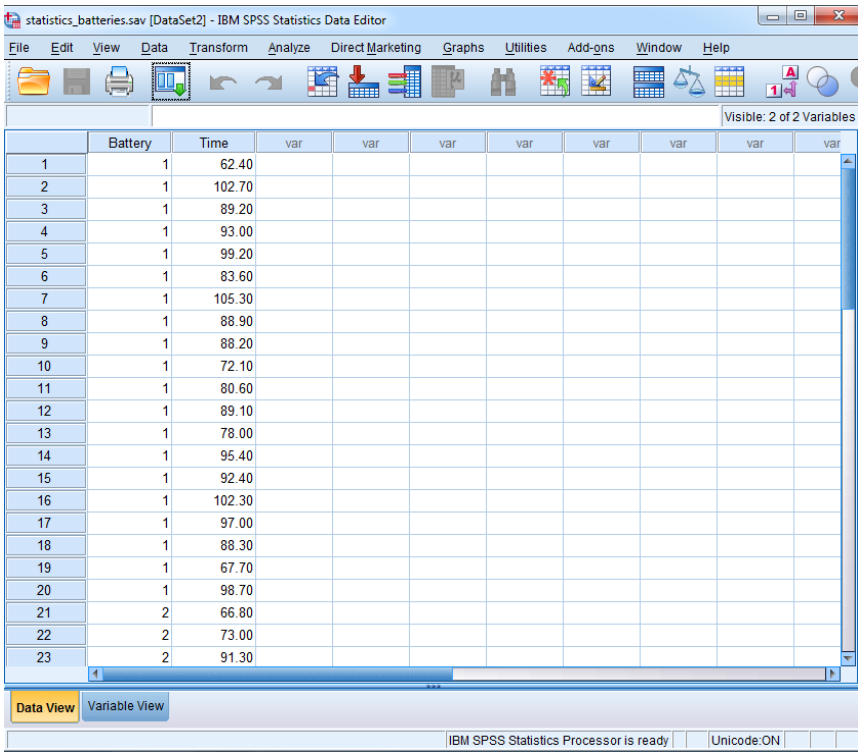
Creating side-by-side boxplots in SPSS

Example: A random sample of 40 cellphones of the same make and model were chosen. Half of the cellphones were randomly selected to have a nickel-cadmium battery put in them and the rest had a nickel-metal hydride battery. The talk time (in minutes) before the batteries needed to be recharged was recorded.

1. Enter the data into SPSS

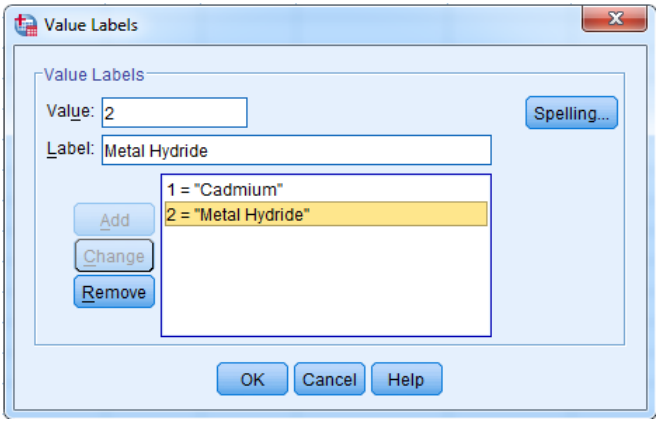
   OR open the Batteries.sav file from **www.tinyURL.com/stats-IS**.

   Use a value of **1** for **Cadmium** and **2** for **Metal Hydride**:



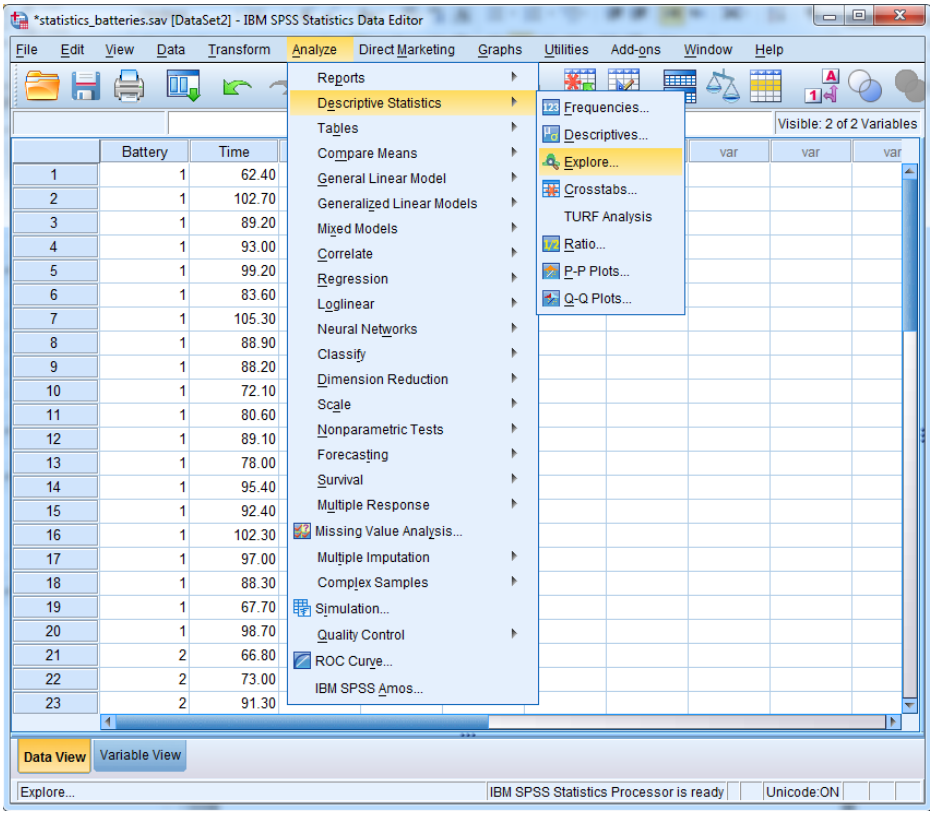
	Battery	Time	var	var	var	var	var	var	var	var
1	1	62.40								
2	1	102.70								
3	1	89.20								
4	1	93.00								
5	1	99.20								
6	1	83.60								
7	1	105.30								
8	1	88.90								
9	1	88.20								
10	1	72.10								
11	1	80.60								
12	1	89.10								
13	1	78.00								
14	1	95.40								
15	1	92.40								
16	1	102.30								
17	1	97.00								
18	1	88.30								
19	1	67.70								
20	1	98.70								
21	2	66.80								
22	2	73.00								
23	2	91.30								

2. Assign labels.
   Label the values:
   Label **1** as **Cadmium** and **2** as **Metal Hydride**.



3. Plot the data using a boxplot.

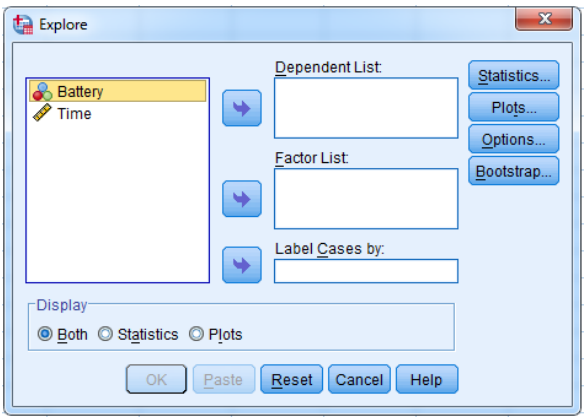
   a. Choose the **Explore** tool:
   Click **Analyze** → **Descriptive Statistics** → **Explore**

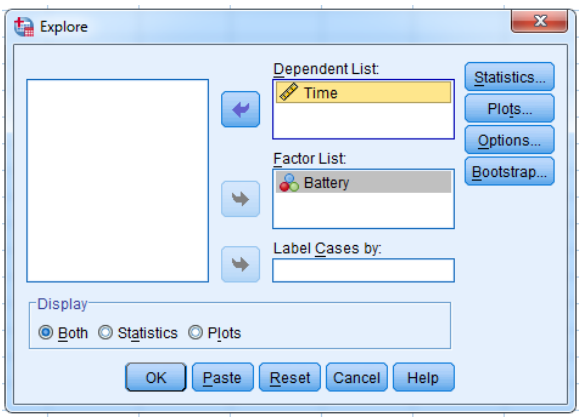


   b. Assign the variables.

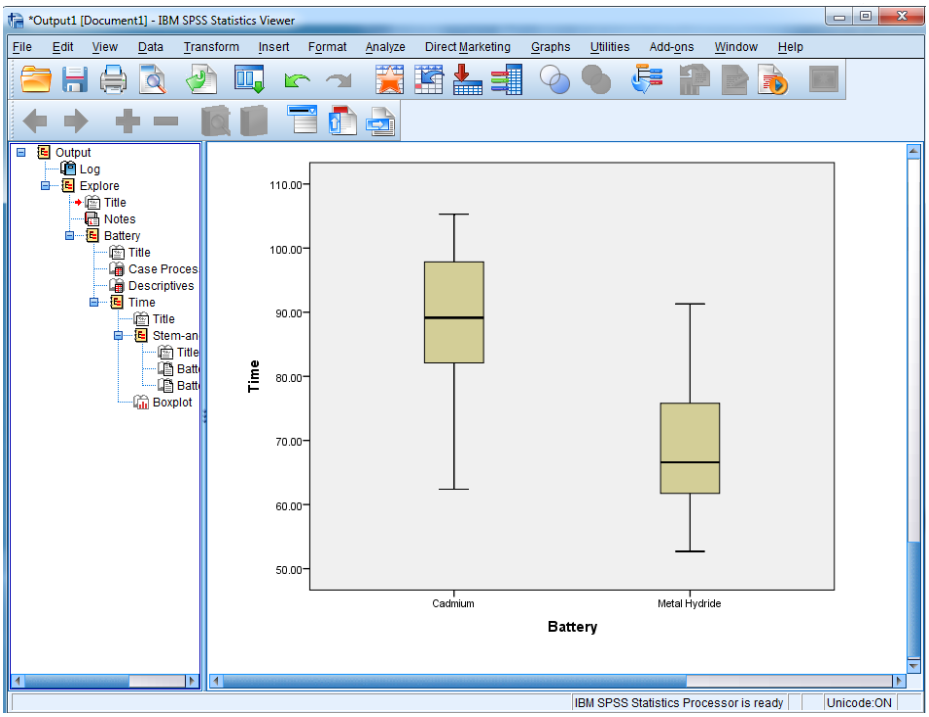
   Quantitative (response) variable → **Dependent List** box.
   Click **Time**.
   Click ▶.

   Qualitative variable (grouping factor) → **Factor List** box.
   Click **Battery**.
   Click ▶.





c. View and interpret the boxplots.



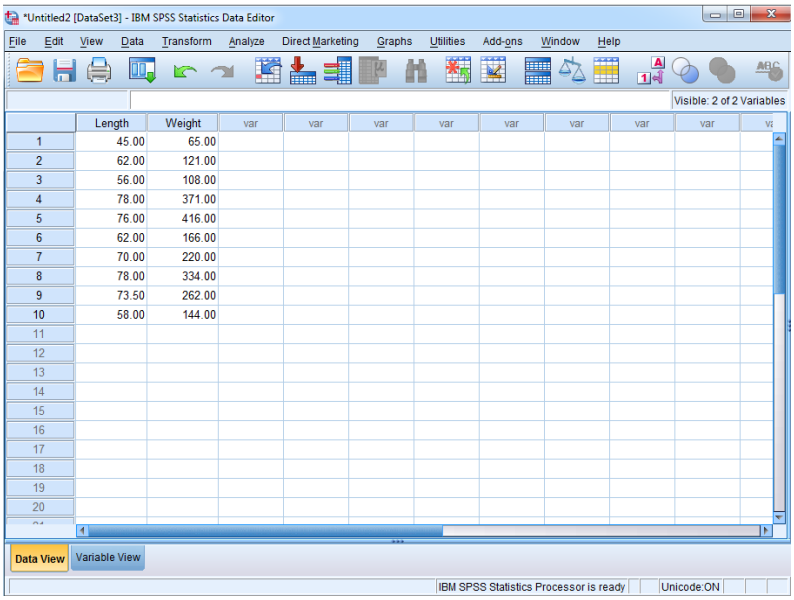
Creating a Scatter Plot in SPSS

Example: Create a scatter-plot of the female coyote length and weight data.

1. Enter the data into SPSS

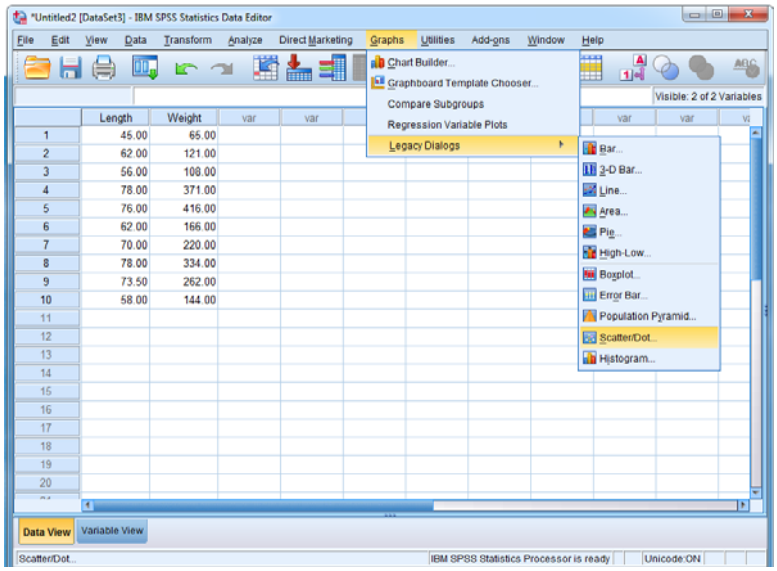
   OR import Length and Weight Data.xls into SPSS after saving it from **www.tinyURL.com/stats-IS**

   OR copy the numerical values from *Excel* to SPSS and enter the variable names **Length** and **Weight** in the **Variable View**.

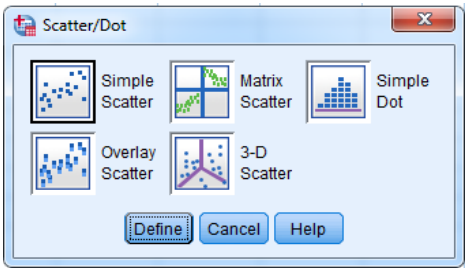


Choose the graph type.

Click **Graphs** → **Legacy Dialogs** → **Scatter/Dot**.



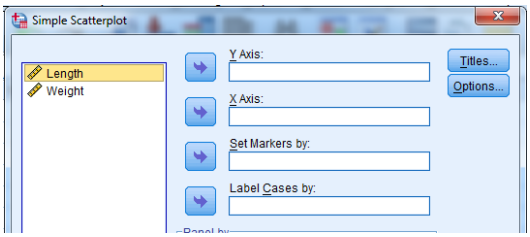
Click **Simple Scatter**.
Click **Define**.

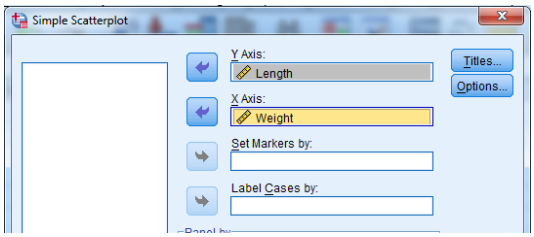


2. Select the relevant variables.

Dependent variable → **Y Axis** box.
Click **Length**. Click the first ▶.

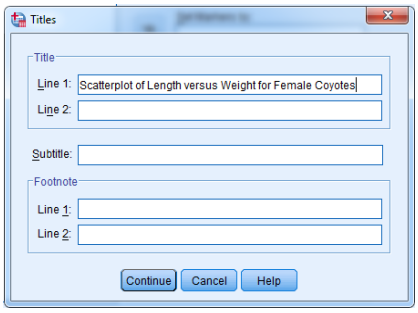
Independent variable → **X Axis** box.
Click **Weight**. Click the second ▶.



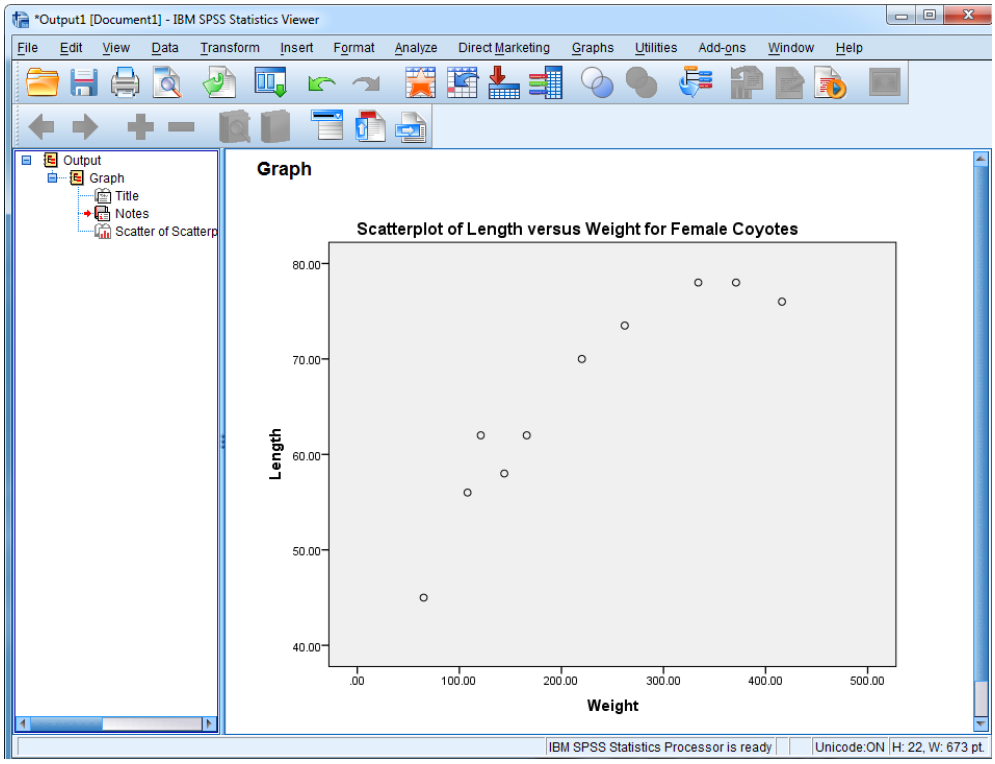


3. Enter titles.

Click **Titles**.
Type **Scatterplot of Length versus Weight** in the **Title Line 1** box.
Click **Continue**. Then click **OK**.

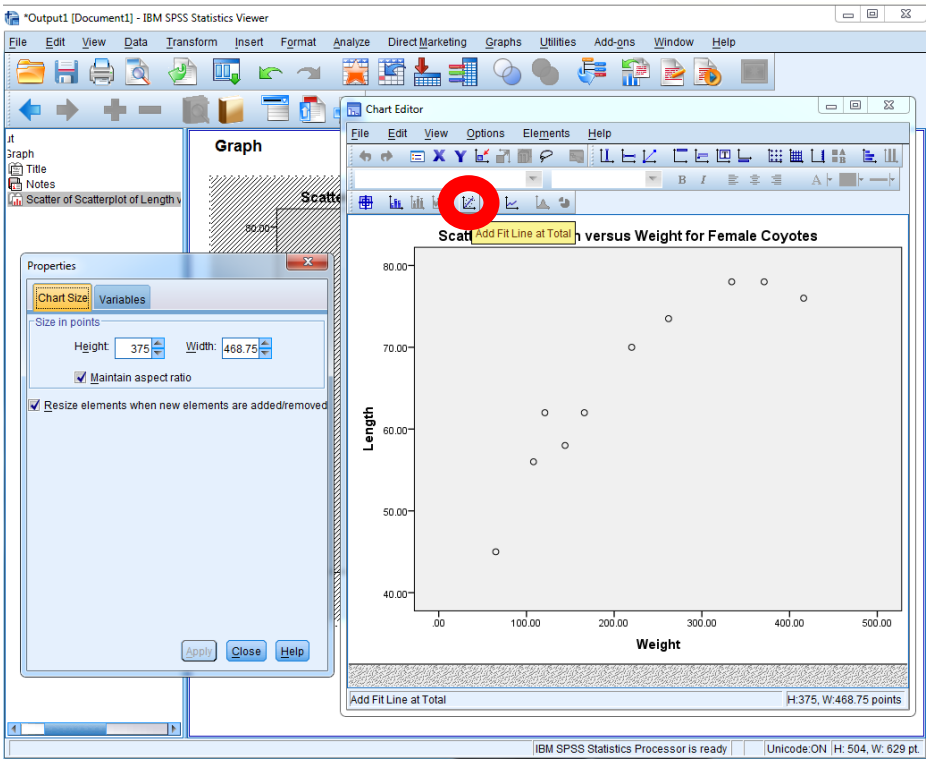


4. The graph appears in the Output Window.

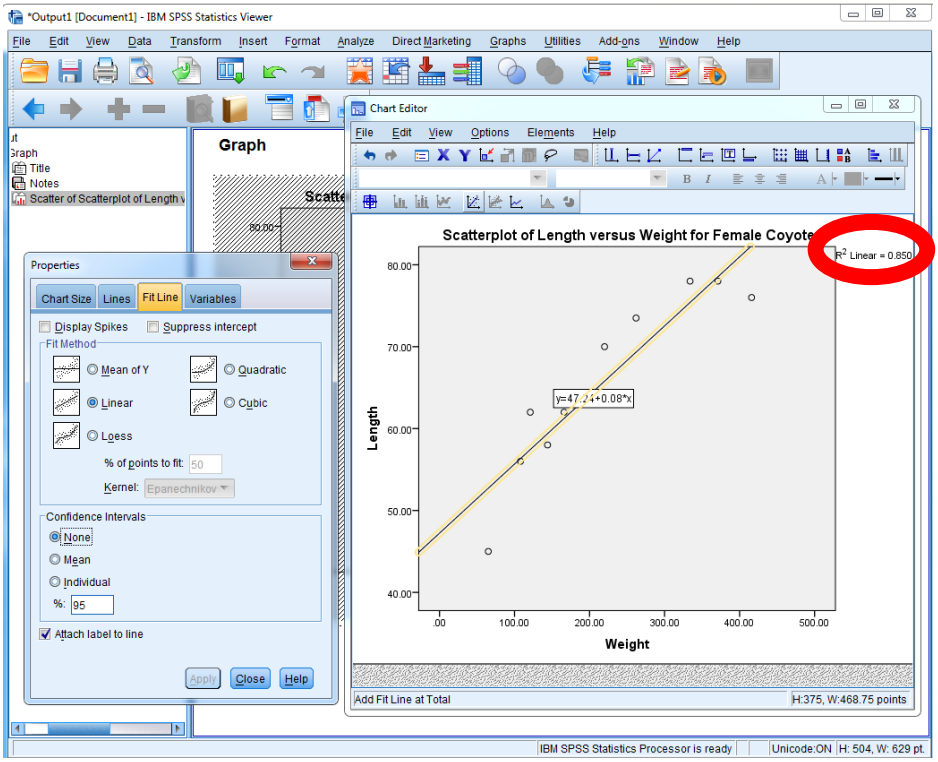


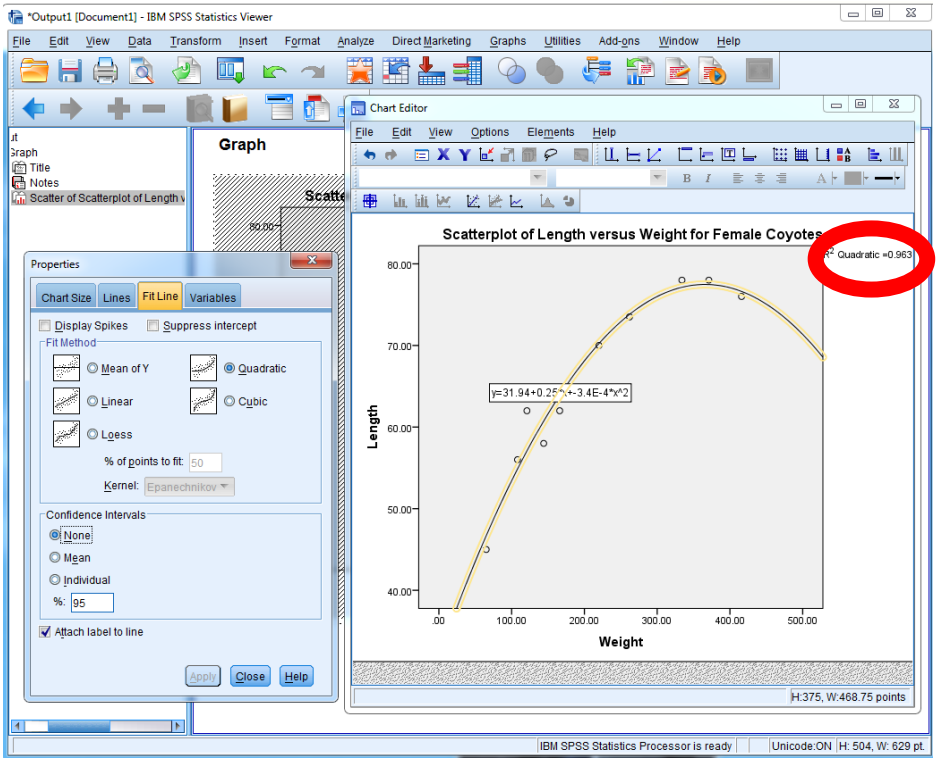
5. To add a trend line to the scatterplot:

   a. Double click on your plot. Click on "Add Fit Line At Total".

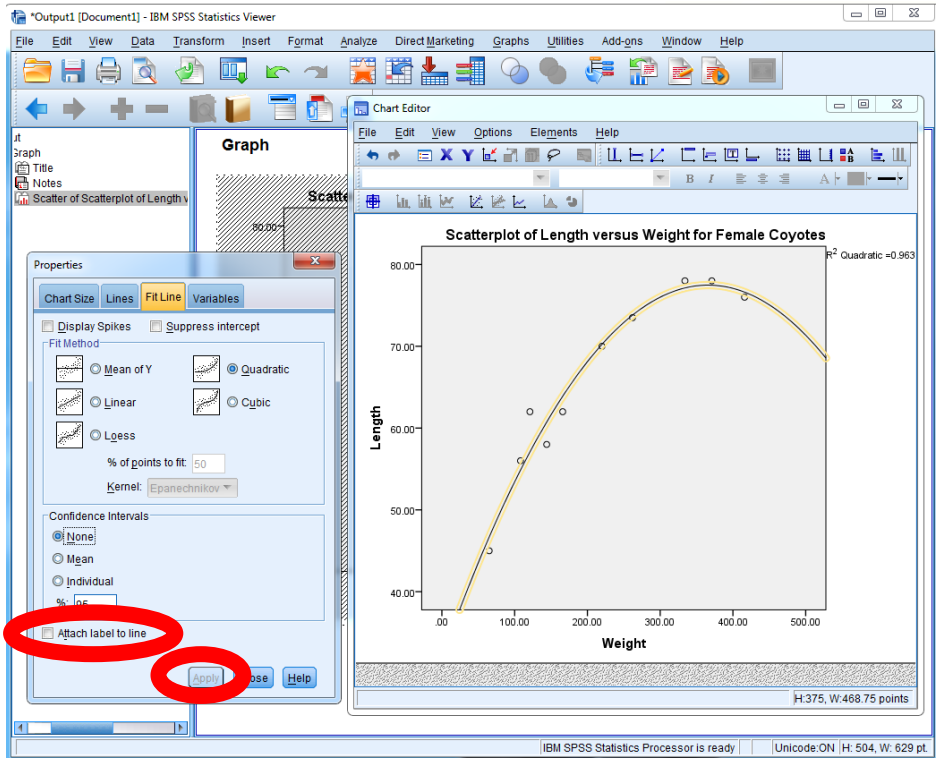


b. Choose the line which gets R^2 closest to 1 (either **Quadratic** or **Cubic** in this case) by clicking on your choice and then clicking **Apply**.





c. Delete any elements you don't want. Uncheck the box next to "Attach label to line" and click **Apply**.



d. Delete any elements you don't want. Click on the R^2 statement and press the **Delete** key then close the graph editor.